DOCUMENT RESUME

ED 202 462 IR 009 334

AUTHOR Lockhart, Kathleen A.; And Others
TITLE Computer-Managed Instruction in the Navy: IV. The Effects of Test Item Format on Learning and Knowledge Retention.
INSTITUTION Navy Personnel Research and Development Center, San Diego, Calif.
REPORT NO NPRDC-TR-81-8
PUB DATE Mar 81
NOTE 33p.; For related documents, see IR 008 994 and IR 009 260.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Computer Managed Instruction; *Constructed Response; Engineering Education; Learning; *Multiple Choice Tests; Outcomes of Education; *Retention (Psychology); Scoring; Student Attitudes; Tables (Data); *Test Format; *Test Items; Time Factors (Learning)

ABSTRACT

The relative effectiveness of multiple-choice (MC) and constructed-response (CR) test formats in computer-managed instruction (CMI) were compared using four test groups of 30 trainees each who were assigned nonsystematically from the basics course at the Propulsion Engineering School, Great Lakes Naval Training Center. Group A took module tests in the standard CR format with answer cues and then converted their answers to MC answer sheets for CMI scoring. Group B took CR tests with answer cues, but the research staff converted the answers. Group C took CR tests but without answer cues, and the staff converted the answers, while Group D took tests in the MC format. Before and after the tests, skills and knowledge were measured to compare factors such as learning, retention, time to complete the course, and attitudes. There were no differences in learning among the groups, but Group C, with the CR questions without cues, had the best retention, but took longer to complete the course and rated their tests as being more difficult than did students in the other groups. The attitude questionnaire and the results of ANOVAs comparing the groups on measures of learning are appended, and five references are provided. (Author/BK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

NPRDC TR 81-8

March 1981

ED202462

COMPUTER-MANAGED INSTRUCTION IN THE NAVY: IV. THE EFFECTS OF TEST ITEM FORMAT ON LEARNING AND KNOWLEDGE RETENTION

Kathleen A. Lockhart
Persis T. Sturges
Nicholas H. Van Matre
Judith Zachai

Reviewed by
James S. McMichael

Released by
James F. Kelly, Jr.
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152

IR009334

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPRDC TR 81-8	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMPUTER-MANAGED INSTRUCTION IN THE NAVY: IV. THE EFFECTS OF TEST ITEM FORMAT ON LEARNING AND KNOWLEDGE RETENTION		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kathleen A. Lockhart Judith Zachai Persis T. Sturges Nicholas H. Van Matre		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 63720N Z1176-PN.01
11. CONTROLLING OFFICE NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		12. REPORT DATE March 1981 13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Test-item type
Multiple-Choice
Retention
Constructed-Response
Learning
Computer-managed Instruction
Propulsion Engineering

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The relative effectiveness of multiple-choice (MC) and constructed-response (CR) test formats in computer-managed instruction (CMI) was compared. Most CMI tests use the familiar MC format with standard answer forms because they can be machine-scored and sent directly into the computer. CR formats, which require students to generate their own written answers to each item, cannot be directly input to the computer, but they permit a more varied range of student responses.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The MC format was compared with three variations of the CR format using four test groups, each consisting of 30 trainees assigned nonsystematically from the basics course at the Propulsion Engineering School, Great Lakes Naval Training Center. No measurable differences were found among the groups in amount of learning. This result implies that the MC format is preferable since it is less costly and is compatible with the current CMI system.

The CR group that was given no prompts or cues as to the possible answers showed better retention of what they had learned. However, this format is least compatible with the CMI system, and was more time consuming for students and staff. Before this CR format could be operationally feasible, costs would have to be controlled significantly--possibly, in part, by developing a CMI capability for automatic processing of CRs.

FOREWORD

This research and development was performed under Work Unit Z1176-PN.01 (Improving the Navy's Computer-managed Training System), as part of a research and development effort aimed at improving the Navy's operational computer-managed instruction (CMI) system. It was sponsored by the Deputy Chief of Naval Operations (OP-01).

This is the fourth of five related but independent reports describing results of the NAVPERSRANDCEN CMI R&D program. Previous reports described the CMI system and the development of the R&D program (NPRDC SR 80-33), the effect of alternate student-to-instructor ratios on student performance and instructor behavior (NPRDC TR 81-6), and the development and evaluation of an automated performance-testing system for teletyping in the Radioman "A" CMI course (NPRDC TR 81-7). This report is concerned with the effects of CMI test-item formats on retention of learning and knowledge. Results of the CMI research will be used by the Chief of Naval Education and Training (CNET), the Chief of Naval Technical Training (CNTT), commanding officers of all the Navy CMI schools, and others concerned with computer-based instruction.

Appreciation is expressed to the instructors and staff of the Basic Course at the Propulsion Engineering School, Great Lakes Naval Training Center, for their extensive help and cooperation during the data collection phase of this study.

JAMES F. KELLY, JR.
Commanding Officer

JAMES J. REGAN
Technical Director

SUMMARY

Problem

The basic course at the Propulsion Engineering (PE) school, Great Lakes Naval Training Station, uses a constructed-response (CR) test format with answer cues. However, since this format is incompatible with the computer-managed instruction (CMI) system, which requires machine-readable, multiple-choice (MC) answers, students must convert answer sheets to numerical form after each test so scores can be entered into the CMI system. This scoring procedure is time consuming, and would be warranted only if there were significant training gains in terms of learning and long-term knowledge retention.

Objective

The objective of this effort was to investigate the effects of different test-item formats upon student learning, knowledge retention, time in training, and attitudes.

Approach

Students were assigned nonsystematically to one of four groups for the duration of the experiment.

1. Group A took module tests in the standard CR format with answer cues and converted answers to an MC answer sheet for CMI scoring.

2. Group B took CR tests with answer cues, but the research staff converted the answers.

3. Group C took CR tests but without answer cues, and the staff converted the answers.

4. Group D took tests in the MC format.

Before and after the tests, skills and knowledge were measured to compare factors such as learning, retention, time to complete the course, and attitudes.

Conclusions

1. There were no measurable differences in learning among the groups.

2. Group D (MC) learned as much as did the three groups using the CR format.

3. Group C, which received CR question without cues, had the best retention; there was no difference in the retention of the other groups.

4. Group C took more time to complete the course, and rated their tests as being more difficult than did students in other groups.

5. Group A, required to convert answer sheets to MC format, took 4.5 hours longer than Group B, whose answer sheets were converted by the staff.

Recommendations

1. The MC format should replace the CR format in PE school tests.

2. If the CR format is continued in use, answer cues should not be provided with the questions. However, consideration should be given to the increased cost of this alternative.

3. The Chief of Technical Training should consider ways to add to CMI capabilities so that it could handle CR test formats, and should conduct cost-analyses of the appropriate alternatives.

CONTENTS

	Page
INTRODUCTION	1
Problem	1
Objective	2
APPROACH	2
Propulsion Engineering School	2
Subjects	2
Materials	4
Pre- and Posttests	4
Comprehensive Test	4
Attitude Questionnaire	4
Variables	4
Independent	4
Dependent	4
Procedure	5
Analysis	6
RESULTS	6
Measures of Learning	6
Mean Number of Items Correct on the Basic Comprehensive Test	6
Mean Gain from Pretest to Posttest	7
Mean Number of Items Correct on the Supplementary Comprehensive Test	7
Measures of Retention	8
Mean Number of Items Correct on the Second Basic Comprehensive Test	8
Amount of Knowledge Loss From the First to the Second Basic Comprehensive Test	8
Mean Number of Items Correct on the Second Supplementary Comprehensive Test	9
Amount of Knowledge Loss From the First to the Second Supplementary Comprehensive Test	10
Time Factors	10
Time Required to Complete the Course	10
Time Required for Taking Tests	11
Time Required for Other Instructional Activities	11
Time Required for Conversion Procedure	12
Number of Tests Taken	12
Conversion and Computer-Scoring Errors	12
Student-Attitude Questionnaire	12
DISCUSSION AND CONCLUSIONS	13
RECOMMENDATIONS	14
REFERENCES	17

Page

APPENDIX A--ATTITUDE QUESTIONNAIRE A-0

APPENDIX B--ANOVA TABLES B-0

DISTRIBUTION LIST

LIST OF TABLES

1. Mean Scores on the Basic Comprehensive Test (First Administration) by Test-Item Format 7

2. Mean Number of Items Correct on Supplementary Comprehensive Test 8

3. Mean Group Responses to Attitude Questionnaire 13

LIST OF FIGURES

1. Examples of test-item type for each experimental group 3

2. Mean retention by group from first to second administration of basic comprehensive test 9

3. Total number of contact hours and the number of hours spent testing for each group 11

INTRODUCTION

Problem

Computer-managed instruction (CMI) is now widely used in much of the Navy's basic technical training schools, since it provides more efficient handling of the large numbers of students in training. The system aids individualized instruction through self-pacing and effective remediation assignments. Testing materials, other than laboratory and performance tests, normally use multiple-choice (MC) or true-false questions as the test-item format, and the answer sheets are machine-scored. As a result, CMI instructors have more time for such critical functions as counseling, tutoring, and monitoring student progress.

The basics course at the Propulsion Engineering (PE) Class "A" School, Great Lakes Naval Training Center, uses a constructed-response (CR) test format. This system is not compatible with the CMI system, since the optical scanner, the student terminal used with the system, precludes the use of such test materials as short-answer or fill-in (CR) items if the tests are to be machine-scored. In spite of this, administrative personnel at the PE school have been reluctant to change to an MC format that could be machine-scored because they believe this format does not provide effective learning and does not enhance retention of skills and knowledge. To obtain some of the advantages of CMI machine-scoring without changing test format, the school developed a conversion procedure to adapt the CR format to CMI requirements. Under this procedure, students convert CR answers to a conventional MC answer sheet. Although this procedure provides some benefits, it is time consuming and involves the risk of inaccurate test scores because of errors during conversion.

In addition to the fact that the conversion procedure is time-consuming, another problem associated with the CR format was perceived that questions its advantage over the MC format. Although the CR format does require the students to write out answers, thereby enhancing learning and retention, approximately 85 percent of the questions are provided with answer cues. It is possible that these cues nullify the advantages the CR format has over the MC format in learning and retention.

Several research studies that relate to these problems have been conducted. For example, Sax and Collet (1968) examined the relation of a mid-term test to a final examination. Half of the students in the study received three MC mid-term tests and the other half, three CR mid-term tests. All of the students were told to expect a CR final. As it turned out, half of each group was given an MC final and the other half, a CR final. Results showed the group that received MC mid-terms performed as well as did the CR group on the CR final, and better than the CR group did on the MC final. The authors noted that these results could be due to the fact that the items in the tests were difficult and required fine discrimination among novel elements. They predicted that, for relatively simple material, the relation observed in the study might not obtain. Unfortunately, they present no guidelines for determining the difficulty of test items to be used in any one course, and the generality of these findings across instructional settings remains to be demonstrated. Their findings do underscore the importance of examining the relation between test-item format, learning, and knowledge retention.

Ulman and Sparzo (1978) examined the relation between test mode and final examination performance in a course taught according to the Personalized System of Instruction (PSI). Half of the students in this study took recognition quizzes (MC, true-false, matching), and half took recall quizzes. At the end of the course, half of the

students in each group were given a recognition type of final examination and the other half, a recall type. Results indicated that type of quiz preparation was not related to student performance on a recognition final examination. However, students who took recognition quizzes scored significantly lower on the recall final examination than did students who took recall quizzes. Further, students in the recognition group took significantly more quizzes to achieve criterion in this mastery-based course than did students in the recall group. Ulman and Sparzo concluded that, if one is concerned with students' ability to recall information rather than simply to choose correct answers, CR tests should be used.

Objective

The objective of this effort was to investigate the effects of different test-item formats upon student learning, knowledge retention, time in training, and attitudes.

This is the fourth of five related but independent reports published describing results of NAVPERSRANDCEN's CMI R&D program. Previous reports described the CMI system and the development of the R&D program itself (Van Matre, 1980), the effect of alternate student-to-instructor ratios on student performance and instructor behavior (Van Matre, Hamovitch, Lockhart, & Squire, 1981), and the development and evaluation of an automated performance testing system for teletyping in the Radioman "A" CMI course (Hamovitch & Van Matre, 1981).

APPROACH

Propulsion Engineering School

The PE School is the Class "A" school for three engineering ratings: Machinist's Mate (MM), Boiler Technician (BT), and Engineman (EN). Before students in these ratings can begin their specialty skill training, they must complete a basics course taught under CMI, which consists of 13 modules of common-core knowledge and skills.[1] The material is self-paced, and the testing is criterion-referenced. Approximately 30 percent of each student's instructional time consists of hands-on training.

Each module in the basics course is divided into lessons. The student works through each lesson and then completes a self-administered lesson test. After the student completes all of the lessons in a module, he takes a module test, which is then computer-scored. If the student achieves 100 percent mastery, he begins the next module; if he does not, he receives either oral remediation from the instructor (if his score is 90% or better) or he is assigned remedial work by the computer (if his score is 70 to 90%). After the student completes all of the 13 modules, he takes a comprehensive test on which he must score 80 percent or better. If he scores below 80 percent, he must retake the test.

Subjects

Subjects were 120 students enrolled in the PE school basics course as of 8 January 1979. These students were randomly assigned to one of four groups.

---

[1]The only difference in requirements is that MMs and ENs must take all four lessons in Module 11, and BTs take only Lesson 1.

1. Students in Group A used the existing PE testing procedure; that is, they constructed their response to the items, 85 percent of which included answer cues. They then converted their answers to MC format for computer-scoring. The conversion sheet listed five answer choices for each item number (the fifth choice was always "None of the above"). The conversion sheet did not include item stems. The student matched his CRs to the MC list and transferred the closest approximation to the computer answer form.

2. Students in Group B received the same CR items and cues as did those in Group A. However, the tests were manually scored, and the computer form was prepared by the experimenters. The frequency of the student conversion errors could be determined by comparing the conversion done by the students in Group A with that done by the staff for Group B.

3. Students in Group C received the same CR items as those in Groups A and B. However, less than 5 percent of the items provided cues. The student constructed his responses, and the experimenters scored the tests and prepared the computer answer sheets.

4. Students in Group D received MC test items, which were constructed by using the stems from the CR items and the five choices from the conversion sheet. Students responded directly on the computer answer form for machine-scoring.

Each group was assigned to a different learning center (LC). As students were dropped from the school or completed the course, new students were assigned to the LCs, so that each group included 30 students throughout the study. The LCs were administered by experienced LC instructors, who were shifted after 4 weeks to place a different instructor in each center.

Figure 1 presents examples of test items for each test format. Two series of all-module tests were constructed for each test format so that students requiring repeated testing took the second test from the alternate series but with the same test format. Each module test had from 25 to 150 questions.

Groups A & B: Constructed Response with Cues

49. To keep from skinning your knuckles when using a wrench, _____ the wrench _____
(pull/push) (toward/away from)

Group C: Constructed Response without Cues

49. To keep from skinning your knuckles when using a wrench, _____ _____ the wrench _____ you.

Group D: Multiple-choice

49. To keep from skinning your knuckles when using a wrench, you should _____ the wrench _____ you.

1. Pull, away from.
2. Push, away from.
3. Pull, toward.
4. Push, toward.
5. None of the above.

Figure 1. Examples of test-item type for each experimental group.

Materials

Pre- and Posttests

The pre- and posttests contained 87 MC items, which were taken from a criterion-referenced test previously developed for the PE course.

Comprehensive Test

The comprehensive test used in this study had four parts. Parts A, B, and C comprised 150 items, half MC and half CR, which were taken directly from the Series I Comprehensive Test in use at the PE school. The number of questions in the MC and CR formats was equated as nearly as possible for each module, for a total of 75 MC items and 75 CR items. Part D comprised 32 CR items from regular PE tests with cues removed for this experiment. Scores on Part D were not used in computing course grades, although students were not informed of this. Hereinafter, Parts A, B, and C of the comprehensive test will be referred to as the basic comprehensive test and part D, as the supplementary comprehensive test.[2]

Two forms of the basic comprehensive test--Forms A and B--were prepared to counterbalance the type of item and the specific questions, one the mirror image of the other. On both forms, about 85 percent of the CR items presented cues.

Attitude Questionnaire

The attitude questionnaires (see Appendix A) included six items concerning the course and testing procedures.

Variables

Independent

Independent variables consisted of three aspects of test item format in the module tests currently in use at the PE school basics course: availability of cues, construction of responses, and conversion of answers. These aspects were systematically varied to compare:

1. Test items that require the student to write his own response (CR) with those that require the student to select one of five choices (MC).

2. Test items that include cues, such as parts lists, with those that do not.

3. Test items that involve the conversion procedure with those that do not.

Dependent

Dependent variables consisted of student attitudes (as measured by responses to the attitude questionnaire) and three aspects of student performance:

---

[2]Since BTs were not required to take Lessons 2, 3, and 4 in Module 11 (see Note 1), material from these lessons was not included in the tests.

1. Learning, as measured by (a) mean number of items correct on the basic and supplementary comprehensive tests and (b) mean gain in score from the pretest to the posttest.

2. Knowledge retention, as measured by (a) mean number of items correct on the basic and supplementary comprehensive tests and (b) mean loss in scores between the first and second administrations of basic and supplementary comprehensive tests.

3. Time factors in the course, including time required to take module tests, to convert answers to MC format, and to complete the course.

Procedure

Students took the pretest before checking into the course on the computer. They were told that the pretest score did not count on their Navy record but that it was important to the research. They were urged to do their best, although they were not expected to know the material. The general administrative procedures for testing currently in use at the PE school were followed (no talking, no papers, etc.).

In taking the various tests, students in all groups (1) brought the computer print-outs directing them to take a test to the test center where they received the appropriate test form and answer sheets, (2) time-stamped answer sheets at the start and end of the test, and (3) returned tests to the experimenters, who graded them and reported the scores to the appropriate LC instructor.

For each group, the method of obtaining the computer read-out with feedback differed slightly:

1. Students in Group A used a conversion sheet to transfer the answers to the computer answer form, time-stamped the answer sheet again when they completed the conversion procedure, and put the answer form through the computer's optical scanner (OPSCAN).

2. Students in Group B and Group C returned to their learning carrels and waited for the experimenter to score the test and prepare the computer answer form before putting the answer form through the OPSCAN.

3. Students in Group D simply put their computer answer forms through the OPSCAN.

All students in all groups (1) returned the answer form to the experimenter who recorded the score from the computer read-out, and (2) took the computer read-out with feedback to the LC instructor.

The comprehensive test was administered in the same way to students in all LCs. Half of the students in each group received Form A of the basic comprehensive test and half, Form B. Following the comprehensive test, students took the posttest, time-stamping it at start and finish.

Comprehensive tests were scored by two independent scorers, and differences were reconciled by a subject-matter expert. Scoring of the pre- and posttests was spot-checked, and no errors were detected. Also, for Group A (conversion group), fill-in

answer sheets were scored by hand to check on deviations resulting from the conversion procedure. For Group D (MC), all module tests were scored by hand to check for errors in computer scoring.

Two weeks after students had completed the course, they returned and took a second comprehensive test. They were told that this test score did not go on their Navy records but that it was very important to the research, and they were urged to do their best.

At the completion of the course, the students anonymously answered the attitude questionnaire about the course and testing procedures.

Analysis

Analyses of variance (ANOVAs) were used to compare the four groups on measures of learning and retention and on time factors. When appropriate, up to three a priori planned orthogonal comparisons were made. These comparisons involved:

1. Group A versus Group B to test for effect of conversion (cued CR test, with and without conversion).

2. Groups A and B versus Group D to test for effects of test format (CR tests with cues versus MC tests with cues).

3. Groups A, B, and D versus Group C to test for effects of tests with cues versus tests without cues.

ANOVA Tables are provided in Appendix B.

RESULTS

Measures of Learning

Mean Number of Items Correct on the Basic Comprehensive Test

The two forms of the basic comprehensive test--Forms A and B--differed only on which items were MC and which were CR. A preliminary ANOVA comparing these two forms across the four test-format groups indicated no significant differences (Table B-1). Consequently, results from Forms A and B were combined for the remaining analyses.

Table 1 provides group mean scores obtained on the 75 MC and the 75 CR items in the first administration of the basic comprehensive test. These means were analyzed by an ANOVA with one between-group variable--test format groups--and one within-subject variable--type of item--and no significant effects were found (Table B-2). The four groups did not differ significantly on their overall score or on the scores for either the MC or CR (with cues) items on this test.

Table 1

Mean Scores on the Basic Comprehensive Test
(First Administration) by Test-Item Format

Group	Item Format	
	MC (N = 75)	CR (N = 75)
A	65.8	65.5
B	65.3	64.4
C	66.1	66.3
D	67.3	65.5

Mean Gain from Pretest to Posttest

The simple ANOVA used to compare the performance of the four groups on the pretest revealed no significant differences among the groups, indicating that the entry-level knowledge of the four groups was equal or similar (Table B-3). The gain from the pretest to the posttest was analyzed by an ANOVA with one between-group variable--test-format groups--and one within-group variable--time of test (pretest or posttest) (Table B-4). The overall mean of the posttest was significantly greater than the mean of the pretest (71.18 vs. 36.85)--$F(1,116) = 2168.96$, $p < .01$. However, there appeared to be no interaction between pretest and posttest scores and the gain from pretest to posttest scores was not significant.

Mean Number of Items Correct on the Supplementary Comprehensive Test

Table 2 presents the mean scores for the four groups on the first and second administration of the supplementary comprehensive test. The mean numbers of items correct on the first administration of the test were analyzed by an ANOVA with one between-group variable--test format groups (Table B-5). Results showed that the groups differed significantly--$F(3,116) = 4.63$, $p < .01$. The mean score for Group C (CR tests without cues) was significantly greater than the combined mean score for the three groups taking tests without cues--$F(1,116) = 4.63$, $p < .01$. There were no significant differences between the other two comparisons of mean scores. These results indicate that practice in responding to CR items with no cues improves performance on this type of item.

Table 2

Mean Number of Items Correct on Supplementary Comprehensive Test

Group	First Administration	Second Administration
A	20.7	21.2
B	21.7	19.8
C	23.9	22.5
D	22.4	22.0

Note. Based on a total of 32 CR items.

Measures of Retention

Mean Number of Items Correct on the Second Basic Comprehensive Test

Group mean scores obtained on the 75 MC and the 75 CR items during the second administration of the basic comprehensive test were computed. These data were then analyzed by an ANOVA with one between-group variable--test-format groups--and one within-subject variable--item type (Table B-6). Results showed that the overall mean for MC items was significantly higher than the mean for CR items (64.84 vs. 63.45)--$F(1,116) = 12.38$, $p < .01$. Test format had no effect on overall performance or on performance on MC or CR items with cues.

Amount of Knowledge Loss From the First To the Second Basic Comprehensive Test

For each test group, mean scores were computed on three basic measures (1) the number correct in the total 150 items, (2) the number correct in the 75 MC items, and (3) the number correct in the 75 CR items with cues. An ANOVA was conducted on each of these sets of data with one between-group variable--test-format groups--and one within-subject variable--time of test (Table B-7).

Results of the analysis of the total score for each basic comprehensive test showed that, as would be expected, the overall scores were significantly lower on the second comprehensive test--$F(1,116) = 65.67$, $p < .01$. As shown in Figure 2, however, the loss for Group C (CR without cues) was less than the loss for the combined means of the three other groups (MC or CR with cues)--$F(1,116) = 5.36$, $p < .05$.

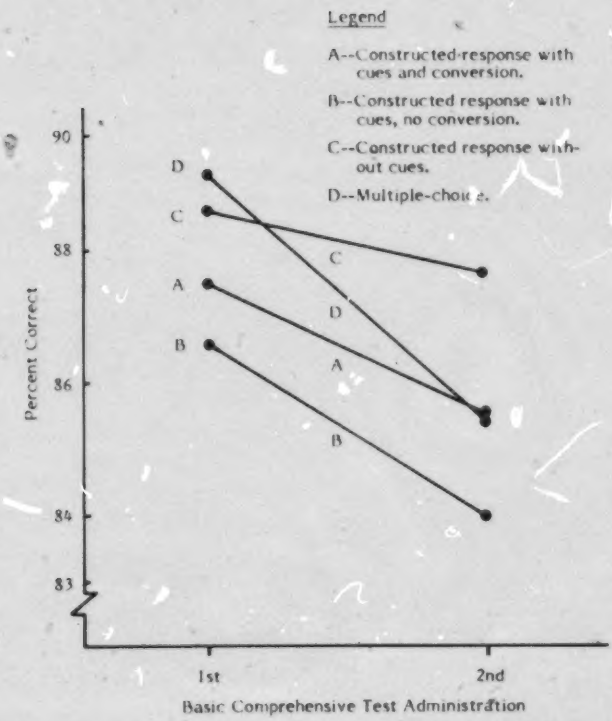


Figure 2. Mean retention by group from first to second administration of basic comprehensive test.

Separate analyses of MC and CR items with cues indicated that the only significant effect was time of test. The scores on the first basic comprehensive test were significantly higher than the scores on the second basic comprehensive test for both MC items--$F(1,116) = 21.51$, $p < .01$--and CR items with cues--$F(1,116) = 43.45$, $p < .01$. As expected, there was a significant loss over the 2-week interval for scores on both types of items, although these losses did not differ for the four groups.

Mean Number of Items Correct on the Second Supplementary Comprehensive Test

Group mean scores obtained on the second supplementary comprehensive test (Table 2) were analyzed by an ANOVA with one between-group variable--test-format groups (Table B-8). Although groups differed significantly ($F(3,116) = 3.00$, $p < .05$), the three *a priori* planned orthogonal comparisons failed to reach significance and did not explain the effect.

Amount of Knowledge Loss From the First to the Second Supplementary Comprehensive Test

To analyze the amount of knowledge loss from the first to the second supplementary comprehensive test, an ANOVA was conducted with one between-group variable--test-format group--and one within-subject variable--time of test (Table B-9). Results showed a significant loss in the number correct over the 2-week interval--$F(1,116) = 9.72$, $p < .01$. This loss differed for the test-format groups--$F(3,116) = 4.34$, $p < .01$.

The interaction of test-format groups and time of test on the supplementary comprehensive test scores was analyzed by the three a priori planned orthogonal comparisons. A comparison of Group A (conversion) and Group B (nonconversion) for CR tests with cues showed that the nonconversion group lost significantly more knowledge than the conversion group ($F(1, 116) = 11.09$, $p < .01$). Since the conversion procedure made a difference, the second comparison of tests with cues was conducted between the two test formats (CR and MC) but did not include the conversion group (Group A). Again, results were significant--$F(1,116) = 4.47$, $p < .05$--with Group B (CR format) losing more than Group D (MC format). The final comparison, between Group B (with cues) and Group C (without cues), did not include the conversion or the MC groups. The results of the comparison were not significant.

Time Factors

Time Required to Complete the Course

The mean number of training contact hours was obtained for two of the PE school LCs that were operating at the same time as those in the study but not involved in the research. These data were computed using all students in each LC and were reported as overall means: LC 1 = 134 hours, and LC 2 = 104 hours.

The mean number of training contact hours for the groups involved in the study were: Group A = 119.64 hours, Group B = 133.30 hours, Group C = 164.41 hours, and Group D = 99.69 hours. Because of the large difference between the time required by Group C (CR, no cues and no conversion) and the other groups, a simple ANOVA was performed between the mean contact hours for this group and those for Group B (CR with cues and no conversion) (Table B-10). Group B was chosen because it was most similar to Group C in testing conditions and had the next highest mean score. Results showed that the average amount of time spent in the course was significantly greater for Group C than for Group B--$F(1,58) = 7.91$, $p < .01$. Assuming equal variance in all groups, it can be inferred that the average amount of time spent by Group C in the course was also significantly greater than that for the other groups.

The total amount of time each student spent taking tests was computed from the time-stamped answer sheets. The time required for the conversion procedures was not included in testing time for students in Group A. The number of contact hours was then partitioned into (1) the time spent testing, and (2) the time spent in other instructional activities (e.g., studying material and job-performance tasks). Figure 3 portrays the mean times for the two categories by group. Means for each time measure were derived by a simple ANOVA with one between-group variable--test-format group.

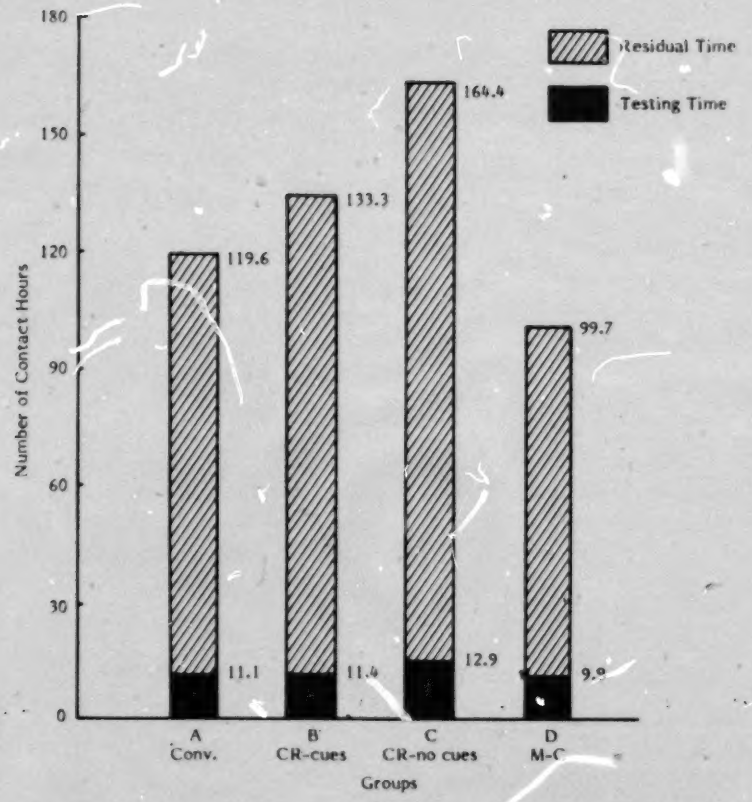


Figure 3. Total number of contact hours and the number of hours spent testing for each group.

Time Required for Taking Tests

The ANOVA performed to compare the four groups on the mean number of hours spent taking tests showed that they differed significantly--$F(3,116) = 3.85$, $p < .05$ (Table B-11). There was no difference between Group A (conversion) and Groups B, C, and D (nonconversion), or between Group D (MC) and Groups A and B (CR with cues). However, Group C (no cues) spent a significantly greater time taking tests than did the combined Groups A, B, and D (cues)--$F(1,116 = 8.46$, $p < .01$. It should be noted that the maximum actual difference between mean test times is between Group C and D, and the mean test-time difference is 3.0 hours.

Time Required for Other Instructional Activities

The ANOVA performed to compare the four groups on the mean number of hours spent in other instructional activities such as studying and performing job tasks also showed a significant difference--$F(3.116) = 16.30$, $p < .01$ (Table B-11). For Group A (conversion), this time included the conversion procedure. There was no difference in the

mean time spent on other activities between Group A (conversion) and Group B (non-conversion).

Group D (MC) spent significantly less time on other instructional activities than did Groups A and B (CR groups with cues)--$F(1,116) = 10.42$, $p < .01$. As a consequence, the final comparison of groups with and without cues did not include Group D. Group C (no cues) spent significantly more time in other acitvities than did Groups A and B (CR groups with cues)--$F(1,116) = 21.20$, $p < .01$.

Time Required for Conversion Procedure

The time spent by Group A in converting the answers averaged 4.27 hours for each student and added significantly to the total time to complete the course (119.6 hours). Group A students took an average of 15.86 module tests.

Number of Tests Taken

One factor contributing to the total time was the number of tests taken. The mean numbers of module tests taken (including retakes, and excluding Module Test 11, Lessons 2, 3, and 4) computed for Groups A, B, C, and D were 15.86, 17.55, 17.81, and 16.38 respectively. These means were analyzed by an ANOVA with one between-group variable--test-format group (Table B-12). Results showed that the groups differed significantly--$F(3,116) = 2.93$, $p < .05$.

Group B (CR, nonconversion) took more tests than did Group A (CR, conversion)--$F(1,116 = 4.97$, $p < .01$. There was no significant difference between Group D (MC) and Group B in the number of tests taken. There was little difference in the average number of tests taken by Group C (no cues) and Groups B and D (cues), although Group C took significantly longer to complete the course.

Conversion and Computer-Scoring Errors

Both CR and MC conversions were hand-scored to assess scoring accuracy. Students in Group A gained an average of 1.78 points per test and lost an average of 1.66 points per test through errors in the conversion procedure. Individual gains ranged from zero to 2.83 points per test; and losses, from zero to 1.93 points per test. These scoring inaccuracies were not large enough either to help or hinder the student. For this study, the maximum number of students tested at one time was 30, with two experimenters and one petty officer proctoring the exams. Greater direct supervision than in the regular testing room may have reduced errors or cheating in the experimental groups.

The computer scoring of the MC tests for Group D was judged as highly accurate by the researchers. Students gained an average of only .03 points per test and lost an average of .11 points per test due to errors in computer scoring.

Student-Attitude Questionnaire

Table 3, which provides mean group responses to the attitude questionnaire, shows that the four groups did not differ on the first three items, which concerned CMI in general, the module books used to present the material, and the tests used to assess knowledge. However, Group A (conversion) was less satisfied about the way tests were given (Item 4) than were the other groups. Most Group A students cited the conversion

procedure as the source of their dislike. As to the difficulty of the tests (Item 5), Group C (CR without cues) said the tests were more difficult than did the other three groups. Finally, the groups differed greatly as to the degree to which they felt their learning supervisor had helped them. Groups B and C felt they had the most help, followed by Group D and Group A.

Table 3

Mean Group Responses to Attitude Questionnaire

Item	A	B	C	D
1. How did you like the computer-managed instruction, in general?	5	5	5	5
2. How well did the module books present the material?	5	5	5	5
3. How well do you think the tests tested your knowledge?	5	5	5	5
4. What did you think about the way tests were given?	4	5	5	5
5. Do you think the tests were difficult?	3	3	4	3
6. How much do you feel that your learning supervisor helped you?	3	5	5	4

Note. Means are based on responses made on a 6-point scale, where 1 = most negative and 6 = most positive. Anchors of items nos. 1 and 4 were "disliked a lot" and "liked a lot"; nos. 2 and 3, "very poorly" and "very well"; no. 5, "no--very easy" and "yes--very difficult"; and 6, "not at all" and "very much."

DISCUSSION AND CONCLUSIONS

The results of this study do not support those obtained by Sax and Collet (1968). The differences in the reported findings may be due to differences in item difficulty, if Sax and Collet are correct in their hypothesis concerning the relation between appropriate item type and item difficulty. The description of test material outlined by Ulman and Sparzo (1978) unfortunately does not permit this sort of analysis. The differences in findings might also be due to the differences in the course format used in the two studies. Sax and Collet conducted their class as a group-paced lecture course; Ulman and Sparzo's class was self-paced with repeated quizzing until mastery was reached. Thus, the PSI subjects not only received more training on a given test mode (greater number of quizzes), but also attained mastery of material. Certainly, this PSI format bears a closer resemblance to the Navy CMI system than does the former, in that Navy computer-managed technical training also demands frequent quizzing and mastery in a self-paced system. Finally, neither of these two studies measured retention of knowledge, a major concern in Navy technical training.

Conclusions based on the results of this study are listed below:

1. Students learned equally well under the four formats. The increase in learning shown for Group C on the supplementary comprehensive test indicated only that students in this group performed better because they had taken tests without cues before and experience gave them an advantage.

2. Format did not affect the retention score of the second basic comprehensive test, but it did affect the amount of loss over the 2-week period. Group C (no cues) showed less loss on items with cues than did the other groups that had had practice on this item type. This result suggests that retention improves when test items require more than the objectives specify.

3. Tests currently used by the PE school (CR with cues and conversion) produced no better learning and retention than did the MC test on any of the criterion test-item formats. Since the conversion requires 4.27 hours per student, much time is lost with no gain in performance.

4. Group C showed better retention (the "real fill-in" group) and took more time to complete the course. This group did not take more tests (including retakes), but spent more time taking tests and performing other activities. Anecdotal data suggests that instructors and students in this group felt they were involved in an unusually relaxed situation without normal pressure. This factor may help explain the increased time in the course.

5. Examination of student attitude data indicated that (a) students taking tests without cues (Group C) rated their tests as being more difficult than did those in the other three groups, (b) students using the conversion procedure (Group A) liked their tests less than did those in the other three groups, and (c) all students generally liked CMI.

6. In assessing and applying the results of the study, consideration must be given to the fact that it was not possible to control the degree of motivation provided by the instructors or the manner in which they provided this motivation. Nor was it possible to assess the quality or quantity of individual tutoring instructors provided or the manner in which they handled oral remediations. These instructor differences could influence the results in a study that measures student performance.

7. No students were sent to "extra study" in an attempt to encourage them to keep up, as is the normal practice at the PE school. Since the procedures of the experiment differed from normal procedures, course completion times could not be projected for the basics course in which CR tests without cues had been incorporated.

RECOMMENDATIONS

1. The MC format should replace the CR format in PE school test.

2. If use of the CR format is continued, answer cues should not be provided with the questions. However, consideration should be given to the increased cost of this alternative.

3. The Chief of Naval Technical Training should consider ways to add to CMI capabilities so that it could handle CR test formats, and should conduct cost-analyses of the appropriate alternatives.

4. Since this study suggests that retention improves when requirements exceed objectives, research should be conducted to determine the best way in which training and tests can be designed to demand more from the students than is required by the specified objectives.

REFERENCES

Hamovitch, M., & Van Matre, N. Computer-managed instruction in the Navy: III. Automated performance testing in the Radioman "A" school (NPRDC TR 81-7). San Diego: Navy Personnel Research and Development Center, March 1981.

Sax, G., & Collet, L. S. An empirical comparison of the effects of recall and multiple-choice tests on student achievement. Journal of Educational Measurement, 1968, 5, 169-173.

Ulman, J. D., & Sparzo, F. J. Differential effects of two types of unit quiz modes on final exam performance. Journal of Personalized Instruction, 1978, 3, 187-191.

Van Matre, N. Computer-managed instruction in the Navy: I. Research background and status (NPRDC SR 80-33). San Diego: Navy Personnel Research and Development Center, September 1980.

Van Matre, N., Hamovitch, M., Lockhart, K. A., & Squire, L. Computer-managed instruction in the Navy: II. A comparison of two student/instructor ratios in CMI learning centers (NPRDC TR 81-6). San Diego: Navy Personnel Research and Development Center, February 1981.

APPENDIX A

ATTITUDE QUESTIONNAIRE

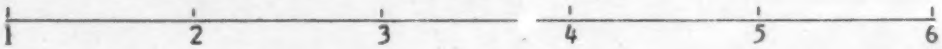
Check one: Learning Center: K____ L____ M____ N____

1. How did you like the computer-managed instruction, in general?

I disliked it a lot — I liked it a lot

Why?_____

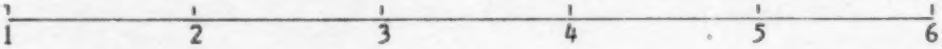
2. How well did the module books present the material?



Very poorly — Very well

Why?_____

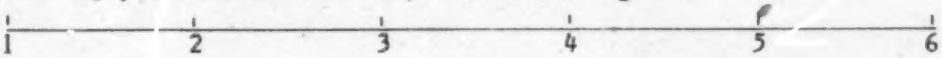
3. How well do you think the tests tested your knowledge?



Very poorly — Very well

Why?_____

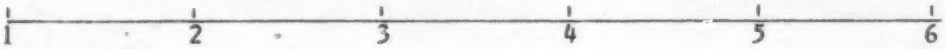
4. What did you think about the way the tests were given?



I disliked it a lot — I liked it a lot

Why?_____

5. Do you think the tests were difficult?



No--very easy — Yes--very difficult

Why?_____

6. How much do you feel that your learning supervisor helped you?

Not at all — Very much

Why?_____

Make any additional comments on any question on the back of this sheet. Also, please make comments or suggestions for improvements--on the back of this sheet.

Thank you!

APPENDIX B

ANOVA TABLES

ANOVA TABLES

Results of ANOVAs Comparing Groups on Measures of Learning

Table B-1

Mean Number of Items Correct on Form A and Form B of the Basic Comprehensive Test

Source	SS	df	MS	F
Form of Test	97.2000	1	97.2000	1.609
Error	7128.0988	118	60.4076	--

Table B-2

Mean Number of Items Correct on the First Administration of the Basic Comprehensive Test

Source	SS	df	MS	F
Group	88.18182	3	29.39394	1.11
Error	3069.66348	116	26.46262	--
Item Type	28.01666	1	28.01666	3.19
Group x Item Type	37.64998	3	12.54999	1.43
Error	1020.33300	116	8.79597	--

Table B-3

Mean Number of Items Correct on the Pretest

Source	SS	df	MS	F
Between Groups	119.0995	3	39.6998	.587
Error	7846.1995	116	67.6397	--

Table B-4

Mean Gain in Score from the Pretest to the Posttest

Source	SS	df	MS	F
Groups	265.71191	3	88.57064	.98
Error (1)	10451.75049	116	90.10130	--
Tests	70692.28613	1	70692.28613	2168.96*
Groups x Tests	14.41260	3	4.80420	.15
Error (2)	3780.74936	116	32.59267	--

*p < .01

Table B-5

Mean Number of Items Correct on the First Administration of the Supplementary Comprehensive Test

Source	SS	df	MS	F
Between Groups	158.4914	3	52.8305	4.633*
Error	1322.8332	116	11.4037	--

*p < .01

Results of ANOVAs Comparing Groups in Measures of Retention

Table B-6

Mean Number of Items Correct on the Second Administration of the Basic Comprehensive Test

Source	SS	df	MS	F
Groups	172.97809	3	57.65936	1.60
Error (1)	4175.41455	116	35.99495	--
Items	116.20415	1	116.20415	12.38*
Groups x Items	32.14580	3	10.71527	1.14
Error (2)	1089.14967	116	9.38922	--

*p < .01

Table B-7

Amount of Loss from the First to the Second Administration of the Basic Comprehensive Test

Source	SS	df	MS	F
MC and CR Items (N = 150)				
Groups	436.67871	3	145.55957	1.27
Error (1)	13333.49048	116	114.94388	--
Test	653.39980	1	653.39980	65.67*
Groups x Tests	67.43329	3	22.47776	2.26
Error (2)	1154.16618	116	9.94971	--
MC Items Only (N = 75)				
Groups	96.47760	3	32.15920	1.08
Error (1)	3462.91711	116	29.85273	--
Test	97.53748	1	97.53748	21.51*
Groups x Tests	20.84582	3	6.94861	1.53
Error (2)	526.11649	116	4.53549	--
CR Items Only (N = 75)				
Groups	198.74902	3	66.24967	1.62
Error (1)	4735.42731	116	40.82265	--
Test	236.01660	1	236.01660	43.45*
Groups x Tests	14.88332	3	4.96111	.91
Error (2)	630.09974	116	5.43189	--

*p < .01

Table B-8

Mean Correct on the Second Administration of the Supplementary Comprehensive Test

Source	SS	df	MS	F
Between Groups	125.9586	3	41.9862	3.003*
Error	1621.6332	116	13.9796	--

*p < .05

Table B-9

Amount of Loss from the First to the Second Administration of the Supplementary Comprehensive Test

Source	SS	df	MS	F
Groups	230.89960	3	76.96653	3.62
Error (1)	2467.03247	116	21.26752	--
Test	40.01665	1	40.01665	9.72*
Groups x Tests	53.54999	3	17.85000	4.34**
Error (2)	477.43317	116	4.11580	--

*p < .01

**p < .05

Results of ANOVAs Comparing Groups on Time Factors

Table B-10

Time Required to Complete Course--Group B Versus Group C

Source	SS	df	MS	F
Between Groups	14520.5801	1	14520.5801	7.916*
Error	106388.2212	58	1834.2797	--

*p <. 01

Table B-11

Times Required by the Four Groups

Source	SS	df	MS	F
Time Required for Taking Tests				
Between Groups	131.1738	3	43.7246	3.851*
Error	1317.2027	116	11.3552	--
Time Required for Other Instructional Activities				
Between Groups	60826.5413	3	20275.5137	16.296*
Error	144333.8506	116	1244.2573	--

*p < .01

Table B-12

Number of Tests Taken by the Four Groups

Source	SS	df	MS	F
Between Groups	82.7582	3	27.5861	2.928*
Error	1092.8333	116	--	--

*p < .05